

2004 年度 修士論文

# サッカー要約映像における Index 情報を利用した自動アナウンス生成

Automatic generation of announcement using  
Index information in Soccer Video summarization

提出日:2005 年 2 月 2 日

指導教授

白井克彦 教授

早稲田大学大学院 理工学研究科 情報・ネットワーク専攻  
Major in Computer Science, Graduate School of Science and Engineering,  
Waseda University

3603U113-1

平林 衛

Mamoru Hirabayashi

# 目 次

第 1 章	序論	6
1.1	研究背景	6
1.2	従来の研究	6
1.3	本プロジェクトの要約映像	7
1.3.1	作成手法	7
1.3.2	問題点	8
1.4	本研究の目的	9
1.5	本論文の構成	9
第 2 章	本研究で目指すアナウンス音声	11
2.1	アナウンス音声の種類	11
2.2	本研究で目指すアナウンス音声	12
第 3 章	システム	13
3.1	研究の概要	13
3.2	システム図	13
第 4 章	分析	15
4.1	サンプルデータ	15
4.2	発話タイミングの調整	16
4.2.1	既存アナウンス音声の構造化	16
4.2.2	発話タイミングのルール	17

4.3	挿入インデックスの選択 . . . . .	18
4.3.1	チーム . . . . .	18
4.3.2	選手 . . . . .	19
4.3.3	イベント . . . . .	19
4.4	語彙変換 . . . . .	20
第5章	実装	22
5.1	プロセス . . . . .	22
5.2	出力結果 . . . . .	23
第6章	評価	28
6.1	定量評価 . . . . .	28
第7章	考察	29
7.1	システム . . . . .	29
7.2	評価 . . . . .	29
第8章	まとめ・今後の課題	30
8.1	まとめ . . . . .	30
8.2	今後の課題 . . . . .	30
	謝辞	33

# 図 目 次

1.1	要約映像作成手法 . . . . .	7
1.2	要約作成の流れ図 . . . . .	8
1.3	アナウンス付加までの流れ図 . . . . .	9
3.1	目標とするシステム . . . . .	14
4.1	a,b,c サンプルデータ . . . . .	15
4.2	イベント分析に使用したラベリングツール . . . . .	16
4.3	既存アナウンス音声の構成 . . . . .	16
4.4	発話タイミング対応図 . . . . .	17
4.5	語彙変換 . . . . .	20
5.1	タイミング対応サンプル 1 . . . . .	26
5.2	タイミング対応サンプル 2 . . . . .	26
5.3	タイミング対応サンプル 3 . . . . .	27
5.4	タイミング対応サンプル 4 . . . . .	27
5.5	タイミング対応サンプル 5 . . . . .	27

# 表 目 次

2.1	既存アナウンス音声の構成 . . . . .	11
2.2	出現インデクスの分類 . . . . .	11
2.3	出現率 . . . . .	12
2.4	インデクス例 . . . . .	12
4.1	発話タイミング対応表 . . . . .	17
4.2	ズレ率 . . . . .	18
4.3	チーム別発話率 . . . . .	18
4.4	出現位置による敵チームの発話率 . . . . .	19
4.5	選手による発話率 . . . . .	19
4.6	イベントによる発話率 . . . . .	20
4.7	語彙リスト . . . . .	21
5.1	インデクス情報 . . . . .	23
5.2	出力アナウンス . . . . .	23
5.3	インデクス情報サンプル 1 . . . . .	23
5.4	インデクス情報サンプル 2 . . . . .	24
5.5	インデクス情報サンプル 3 . . . . .	24
5.6	インデクス情報サンプル 4 . . . . .	24
5.7	インデクス情報サンプル 5 . . . . .	24
5.8	出力アナウンスサンプル 1 . . . . .	25
5.9	出力アナウンスサンプル 2 . . . . .	25

5.10 出力アナウンスサンプル 3 . . . . .	25
5.11 出力アナウンスサンプル 4 . . . . .	26
5.12 出力アナウンスサンプル 5 . . . . .	26
6.1 誤り率 . . . . .	28
6.2 正解率 . . . . .	28

# 第1章 序論

## 1.1 研究背景

これまでの放送の形態では数局の放送局が莫大な数の視聴者に対して一方的に番組を配信してきた。ところが、近年のブロードバンド普及に伴うインターネット上の動画コンテンツの増加、CS や BS 等の新たな TV 放送局の参入、情報の多様化により視聴者が主体的に見る物を選択できるような双方向性型の放送に移り変わろうとしている。また、今後の放送局の増加やインターネット放送などの本格化により映像メディアを扱う機会はますます増加していくと考えられる。

映像メディア、映像コンテンツが増加することは情報が多様化する反面、その複雑性のために内容が把握しきれなくなる恐れがある。今後の映像メディア、映像コンテンツの増加を考えると、見たいものを見たい部分だけ容易に選択できるようにならなければ、映像メディアの増加は人間にとって好ましくないものになりかねない。

このような背景から映像メディアを効率的に利用するためのシステムへのニーズが高まってきている。映像を効率的に利用するための技術に、映像の自動要約がある。短時間で内容を把握するために、映像を要約することは必要不可欠である。

## 1.2 従来の研究

自動要約映像作成手法の一例として、ユーザが過去に見たフレームと、早送りによりスキップしたフレームの特徴量からユーザの好みを学習し、スキップされずに見た映像と類似した映像を繋ぎ合わせることによって要約映像を作成する方法 [1] がある。しかしこれは要約映像作成までに、ユーザーの履歴を取るという労力を要する。また特定のイベントだけを抜き出したい場合、汎用性が無い。

現在、要約映像作成手法として主流になっているのが、インデックスの重み付けによる必

要シーンの取捨選択技法 [2] である. また同時に, これを補完する技術として動画像への自動インデクシング (索引付け) が研究されている. これは図 1.1 のような手法である.

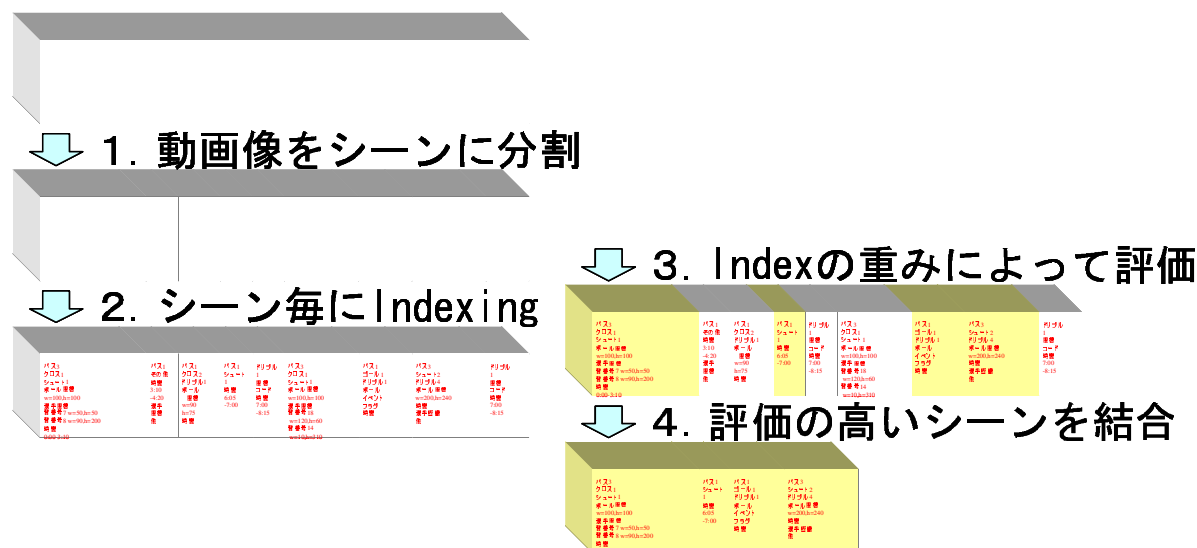


図 1.1: 要約映像作成手法

インデクスとは、動画像の内容を特徴付けるものである。サッカーなどスポーツ映像は特徴のあるイベントが多い。そして、スポーツ映像を題材に様々な自動インデクシングに関する研究が行われている。例えば、サッカーにおいてコーナーポストや選手、ボールの位置を画像から抽出することによりコーナーキックを検出する技法 [3] がある。しかし、動画像解析によるイベント検出は極めて困難な課題である。例えば、画質、カメラワーク等により検出精度が左右されるという問題がある。これらの問題を解決するために別アプローチとして、画像、音声、言語認識など様々なメディアを複数利用した手法 [4] も存在する。

### 1.3 本プロジェクトの要約映像

### 1.3.1 作成手法

本プロジェクトで目標とする要約映像 [5] は，ニュース番組やダイジェスト映像で見られる得点シーンを中心とした一連のプレイの流れにそった要約映像である．作成方法を図 1.2 に示す．

まず，サッカー映像の重要シーンを音声の Power[6] とリプレイシーン [7] により篩いに



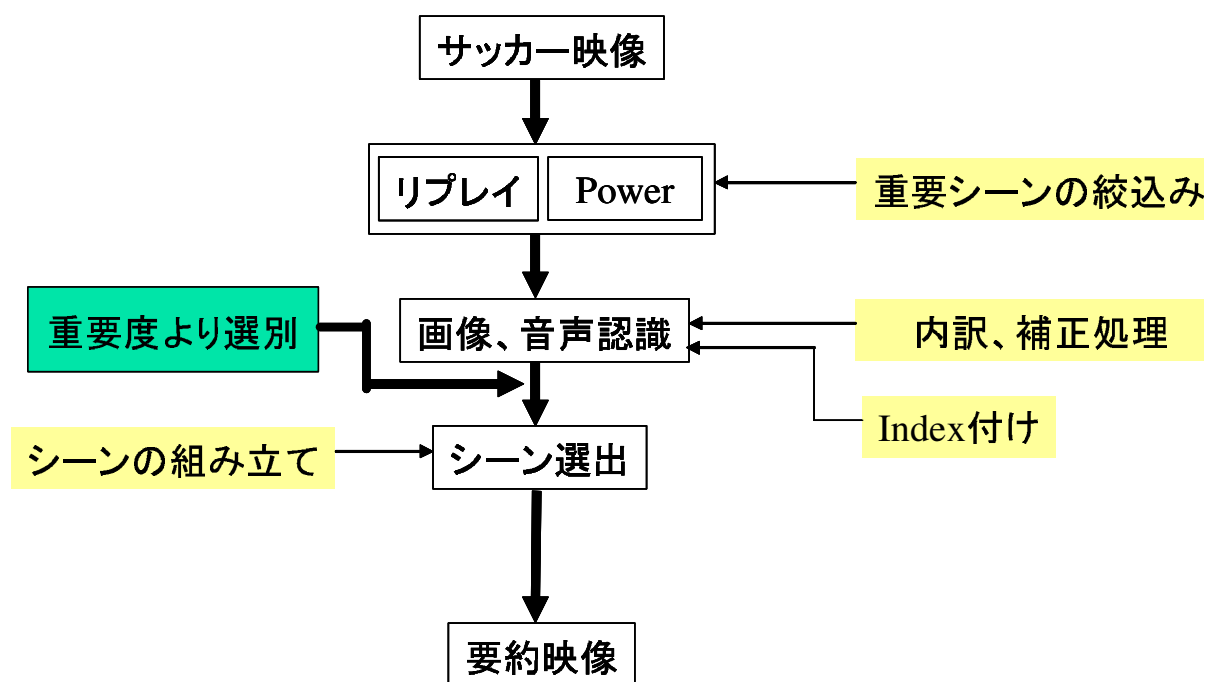


図 1.2: 要約作成の流れ図

かけ，処理にかかる負荷を減少させる．画像認識，音声認識により，画像と音声のズレの補正，インデクス情報の付加を行う．ダイジェスト番組やニュース映像，試合直後のハイライト映像など，すでに人の手によって作成された要約映像をお手本にし，インデクスの重みを決定する．重みの大きい順に映像を組み合わせることにより要約映像が作成される．

### 1.3.2 問題点

ここで出力される要約映像は音声の情報が欠如している．要約映像は時系列上に繋がっていないシーンを組み合わせて作られているため，ユーザには本来与えられるはずである直前の試合状況が送信されない．ユーザの内容理解，内容把握を目的とした要約映像の作成には，新たにアナウンス音声を付加し，映像に対する状況説明を加える必要がある．これによりユーザは試合の状況を瞬時に理解することが可能となる．また，アナウンス音声は映像を盛り上げる演出的効果も発揮する．

## 1.4 本研究の目的

本研究ではサッカー要約映像に付加するアナウンスの自動生成を目的とする。作成方法としては要約映像を作成する際に付加したインデクス情報を再利用し、既存アナウンスを分析することにより生成ルールを作成する。

要約映像作成とアナウンス付加の流れを以下図 1.3 に示す。

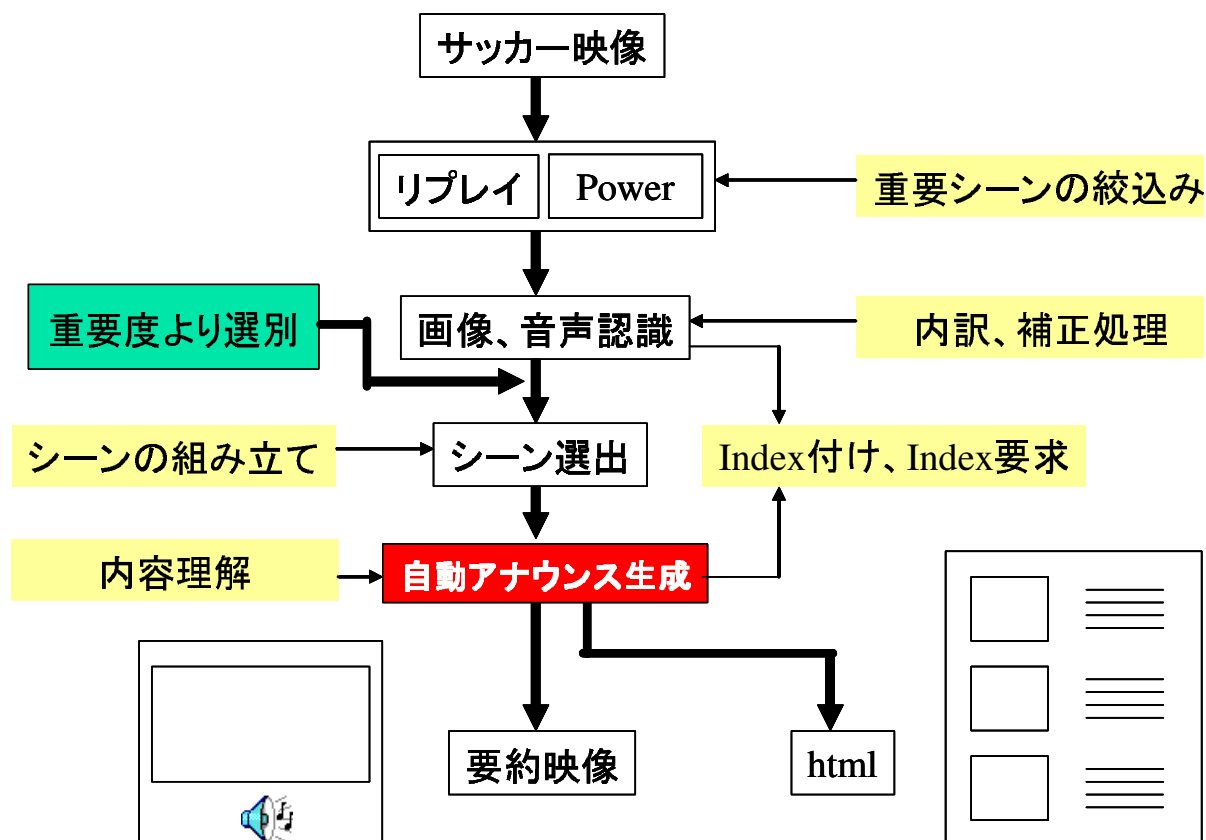


図 1.3: アナウンス付加までの流れ図

選出されたシーンにアナウンスを付加する。アナウンス内容はインデクス情報より取得する。出力形式は音声付要約映像、もしくは映像と説明文の添え書きという HTML 形式のものを想定する。

## 1.5 本論文の構成

本論文の構成は以下 8 章より構成される。

1. 序章 本研究の背景と目的, および本論文の構成を説明する.
2. 本研究で目指すアナウンス音声 本研究で目指すアナウンス音声について説明する.
3. システム 本研究で立案するシステムを説明する.
4. 分析 本研究の分析結果を報告する.
5. 実装 本研究で立案したシステムの実装方法, 出力結果を報告する.
6. 評価 本研究で立案したシステムの評価を報告する.
7. 考察 本研究で立案したシステムについての考察を述べる.
8. まとめ・今後の課題 本研究から得られた結論, 今後の課題について述べる.

## 第2章 本研究で目指すアナウンス音声

### 2.1 アナウンス音声の種類

アナウンス音声には、現在の試合展開、選手の起こしたイベントをそのまま発話する「実況」と、そのイベントに対しどのようなプレイだったかを説明する「解説」が存在する。以下表 2.1 に EURO2004 試合終了後ハイライト 4 試合より得られた既存アナウンス音声の構成 (A:アナウンサー, B:解説者) である。また、以下表 2.2 に出現インデックスの分類、表 2.3 にその出現率を示す。

表 2.1: 既存アナウンス音声の構成

A:実況 + B:プレイ解説	57.70%
A:実況 + B:プレイ解説 + A:戦況	13.50%
A:実況 + B:プレイ解説 + A:プレイ詳細	15.40%
A:実況 + B:プレイ解説 + A:プレイ感想	11.50%
その他	1.90%

表 2.2: 出現インデックスの分類

分類	説明	例
接続詞	文章の繋ぎとなる接続詞	まず、そして、最後に
選手名	アクションを起こした選手の名前	ロナウド、リバウド
イベント	選手が行ったアクション	シュート、ゴール、ドリブル
チーム	出場チーム名	日本、アルゼンチン
時間	要約中の時間	前半 分、後半 分、ロスタイム
場所	フィールド位置	右サイド、左サイド
戦況	試合の流れを表す語彙	チャンス、攻勢
ボール	ボールの状態	こぼれ球
足	蹴り足	右足、左足
点情報	要約中の点、及び点差	1 対 0, 1 点差
ポジション	ポジション情報	ディフェンス
形容詞	形容詞	凄い、鋭い
その他	どれにも当てはまらない語彙	あっと、さあ

表 2.3: 出現率

分類	出現数	出現率
選手名	98	25.26%
イベント	77	19.85%
接続詞	53	13.66%
時間	45	11.60%
チーム	30	7.73%
戦況	19	4.90%
場所	17	4.38%
形容詞	11	2.84%
ボール	8	2.06%
点情報	8	2.06%
足	4	1.03%
ポジション	3	0.77%
その他	15	3.87%
全体	388	100.00%

## 2.2 本研究で目指すアナウンス音声

本研究で目指すアナウンス音声は「実況」である。実況は全シーンにおいて表 2.1 より 98.1%の確率で使用される。またインデクス情報としては、表 2.3 より出現頻度の高い選手名、イベント、時間、チームを使用する。接続語はシーンとシーンの繋ぎ目をスムーズに結ぶために新たに付けられた別ファクターとして考える。

例として、表 2.4 のようなインデクスが振られていた場合、アナウンス音声は「前半 1 分 A チーム、選手 1 から選手 2 のパス、選手 2、選手 2 のシュート!」というようになる。

表 2.4: インデクス例

チーム	選手	イベント	開始時間	終了時間
A	選手 1	パス	0:00:00	0:00:06
A	選手 2	ドリブル	0:00:06	0:00:21
A	選手 2	シュート	0:00:21	0:00:40

## 第3章 システム

### 3.1 研究の概要

アナウンス音声を作成するには以下3つの分析をする必要がある。

- 発話タイミングの調整
- 挿入インデックスの選択
- 語彙変換

アナウンス音声は各イベントに対し、「誰がどうした」と毎回出力されるわけではない。シーン区間の長さによっては抜け落ちる情報、付加される情報が存在する。また、アナウンス音声はインデックスをそのまま出力するわけではなく、適切な語彙へと変換して出力する必要がある。文章として意味の通じない音声は、映像の理解度、またユーザの興奮度を阻害する要因ともなる。そして最後にこれらを全て踏まえた上で発話タイミングを考える必要がある。発話タイミングの逸脱した音声もまた、上記の理由で不適切である。

以上3つは既存アナウンス音声より分析し、特徴を抽出する。

### 3.2 システム図

システムの流れは以下の通りである。

1. インデックス情報入力
2. 発話タイミングの調整
3. 挿入インデックスの選択

4. 語彙変換

5. アナウンス音声出力

これを図 3.1 に示す.

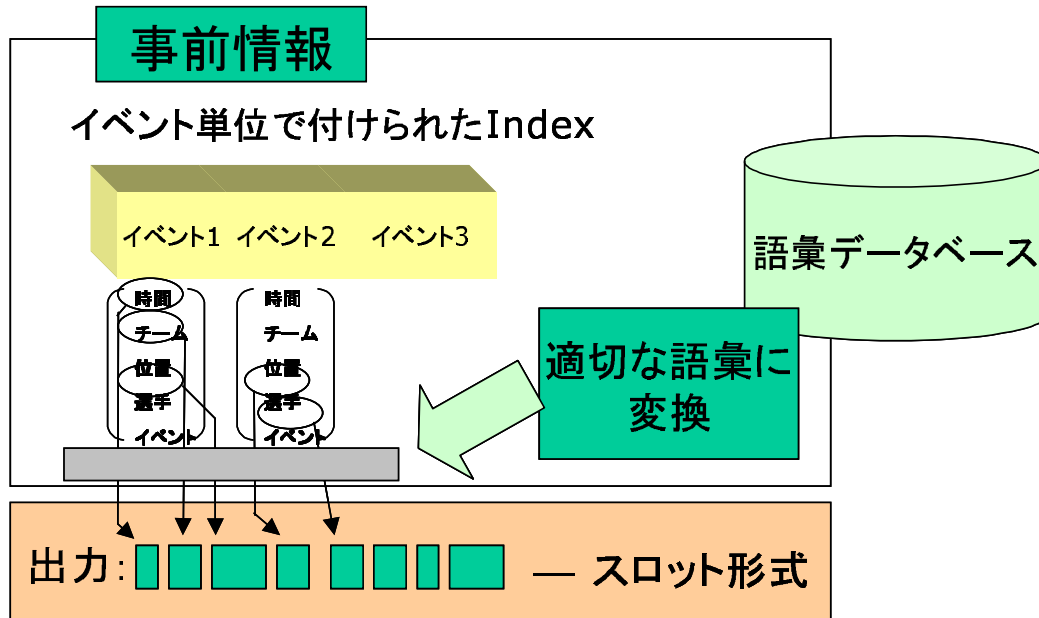


図 3.1: 目標とするシステム

事前情報として、インデクス付き要約映像があるものとする。これはシーンをさらに細分化したイベント単位でインデクスが付けられている。サッカーにおいてイベントとはボール回りにおける選手が起こしたアクションである。インデクスには時間、チーム、位置、選手、イベントが振られている。これより任意のインデクスを決められたスロットに決められたタイミングで挿入する。挿入する際は語彙データベースよりアナウンスに最適な語彙へと変換してから挿入する。

## 第4章 分析

### 4.1 サンプルデータ

アナウンス音声の生成ルールを決めるためには指標となるものが必要である．本研究では現在ダイジェスト番組に付加されている既存のアナウンス音声より発話インデクス，語彙変換，発話タイミングの法則性を分析する．分析サンプルとしては，a) アナウンス書き起こしデータ，b) イベントラベリングデータ，c) 音声発話時間データを使用する．データ a,b,c は以下図 4.1 のような構成である．

前半21分のドイツです。左サイド、シュナイダー。ディフェンダーからボールを奪ってクロスに。シュバインシュタイガーが落としてバラック。

チーム	選手	イベント	開始	終了
B	選手1	パス	0	1
B	選手2	ドリブル	1	2.5
A	選手3	ボールを奪う	2.5	4.5
A	選手3	クロス	4.5	5.7
A	選手4	落とす	5.7	6.8
A	選手5	シュート	6.8	9.2

(a)

(b)

アナウンス	開始	終了
前半21分のドイツです	1.7	3.5
左サイド	3.9	4.5
シュナイダー	4.5	5.2
ディフェンダーからボールを奪ってクロスに	5.2	6.9
シュバインシュタイガーが落として	6.9	7.9
バラック	8.2	9.1

(c)

図 4.1: a,b,c サンプルデータ

これらはEURO2004 試合終了後ハイライトより作成した。(a) アナウンス書き起こしデータは映像に付加しているアナウンス音声を書き起こしたもの、(b) イベントラベリングデータは映像内のボール回りのイベントをチーム、選手、開始時間、終了時間と共にラベリングしたもの、(c) 音声発話時間データはアナウンス音声に発話時間を付加したものである。

また、ラベリングツールとして以下図 4.2 のものを使用した。



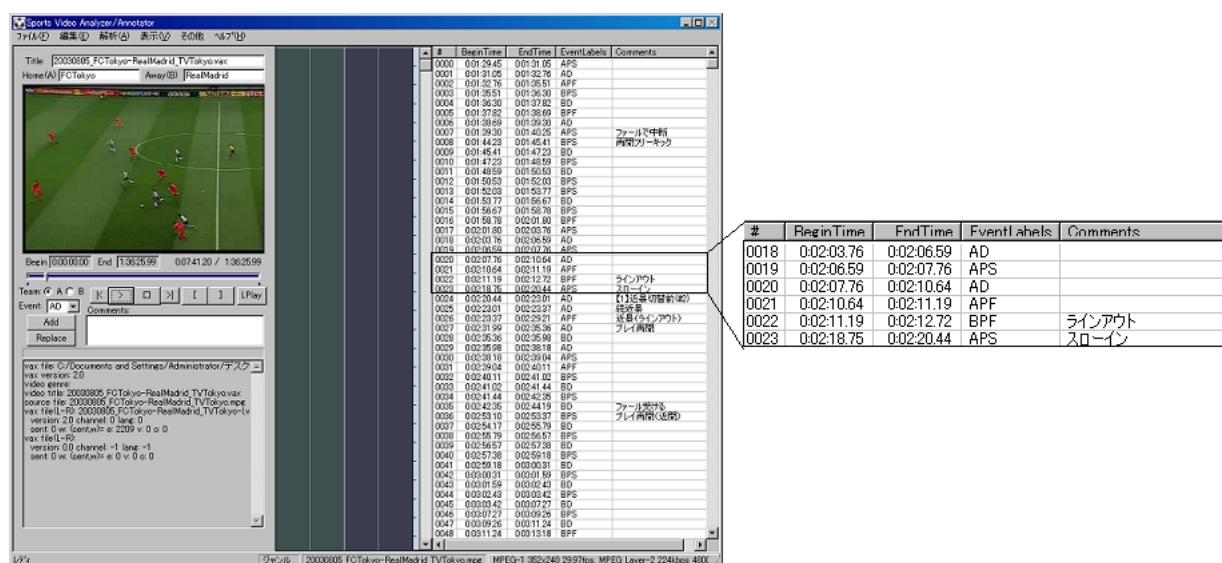


図 4.2: イベント分析に使用したラベリングツール

## 4.2 発話タイミングの調整

### 4.2.1 既存アナウンス音声の構造化

各シーンにおけるアナウンス音声の出現構成は以下図 4.3 のようになる。

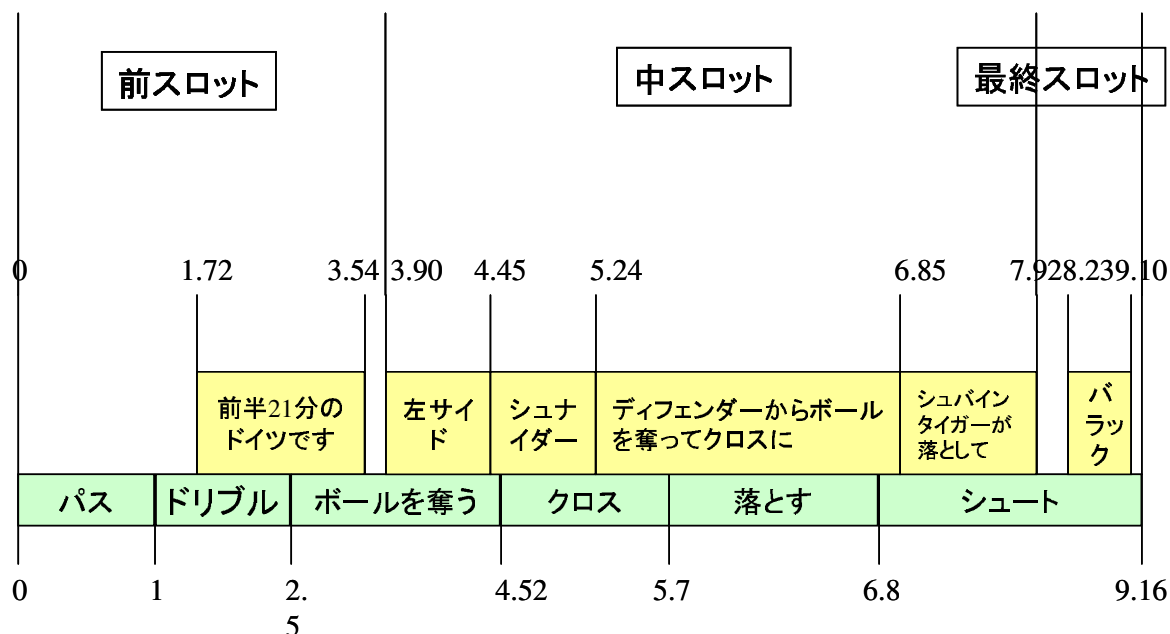


図 4.3: 既存アナウンス音声の構成

上位部分が発話毎に区切られた時間ラベルで，下位部分がイベント毎に区切られた時間ラベルである．これを時系列順に並べている．ここで注目すべきは，図 4.3 に示した通り，発話は前スロット，中スロット，最終スロットと分けることが可能であるということである．主に前スロットでは時間，攻めチームなど試合の状況を説明し，最終スロットではそのシーンにおいて最終イベントがイベントの最後に合うようにタイミングよく発話される．中スロットでは順次行われたイベントが発話されるが，そこにはタイミングのズレなどが観測される．

#### 4.2.2 発話タイミングのルール

本研究では，前述の前スロット，最終スロットには固有の発話区間が存在するとし，既存アナウンス音声よりその平均発話タイミングを算出する．図 4.4 は算出した区間の詳細を現し，表 4.1 はそれに対応する算出結果である．

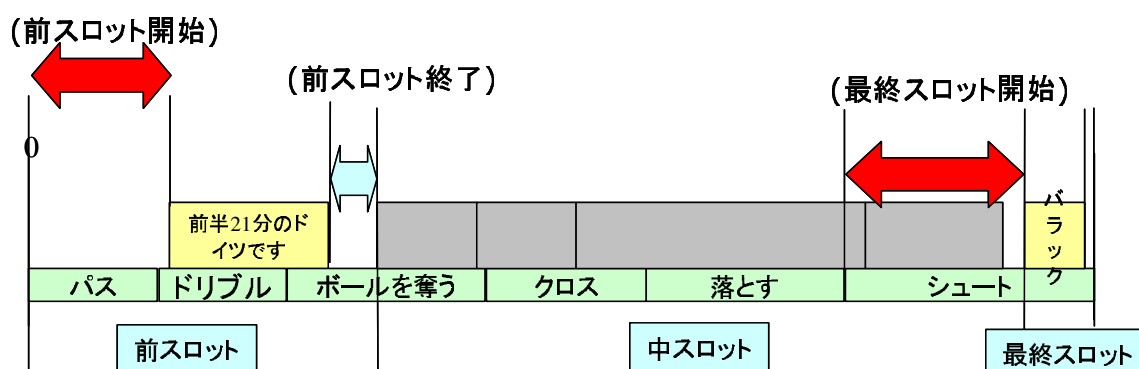


図 4.4: 発話タイミング対応図

表 4.1: 発話タイミング対応表

前スロット開始	前スロット終了	最終スロット開始
0.343 秒	0.843 秒	0.880 秒

前スロット開始は一シーンにおいて最初のイベント開始時間から最初の発話開始時間までの空き時間である．前スロット終了は最初の発話終了時間から次の発話開始時間までの空き時間である．これを算出することにより時間，チーム情報を発話してから実際のイベント発話に入るまでの溜め時間が算出できる．最終スロット開始は最後のイベント

開始時間から最後のイベント発話時間までの空き時間である。

残る中スロットは、前スロット、最終スロットに挟まれた区間のことである。ここで算出出来る法則は、「実際のイベントとアナウンスはズレて行われている」ということである。映像内でアナウンサーがイベントが起こる前にこれから起こることを発話してしまった場合、次の展開を予想するユーザの面白みを削いでしまう可能性が生じる。発話は実際のイベントよりも遅れて発話される。この裏付けとして表 4.2 の結果がある。

表 4.2: ズレ率

ズレイベント数	全イベント数	ズレ率
54	64	84.4%

前スロットから最終スロットまでの空き区間、また上記法則より中スロットの発話区間を算出する。算出した区間には、その区間に収まるよう、優先度の高い順に発話インデクス、変換後の発話語彙を挿入していく。

### 4.3 挿入インデクスの選択

中スロットへ挿入するインデクスの優先度を決定する。インデクスは優先度の高い順に発話区間内に挿入される。優先度の決定には、1) チーム、2) 選手、3) イベントとそれぞれ別に優先度を決定し、先に記したもののほど優先度のルールが上位のものとする。優先度は既存アナウンス音声より分母を実際数、分子をアナウンサーが発話した数とし、発話率が高いもののほど優先度が高いものとする。

#### 4.3.1 チーム

以下表 4.3 はチームによる優先度である。

表 4.3: チーム別発話率

チーム	発話数	選手数	発話率
攻	42	51	82.35%
守	8	16	50.00%

アナウンス音声は攻めチームを基点に構成されている．したがって敵チームの出現数、発話率は低い．また、アナウンスは攻めチームがボールを持った時点で始まる場合が多く、映像にそれ以前のイベントが含まれていた場合、省略されることが多い．以下表 4.4 が敵選手の出現位置による発話率である．

表 4.4: 出現位置による敵チームの発話率

位置	発話数	選手数	発話率
最初	0	3	0.00%
途中	4	6	66.67%
最後	4	7	57.14%
全体	8	16	50.00%

### 4.3.2 選手

次に選手名の挿入優先度の決定である．以下表 4.5 に発話率を示す．

表 4.5: 選手による発話率

	発話数	全選手数	発話率
1人目	14	22	63.64%
2人目	17	18	94.44%
3人目以降	11	11	100.00%

選手名はボールを持っている選手が切り替わった時に発話される．しかし、シーンが始まって最初にボールを持っていた選手の名前は省略される場合がある．それ以降の選手は高い確率で発話される．

### 4.3.3 イベント

最後に各イベント毎の優先度の決定をする．イベント数はそのイベントを表す全語彙をカウントし、全く発話されなかったもののみを省略イベントとする．以下表 4.6 に示す．

ドリブル、シュート、パスなどは選手名を連呼するだけのことが多いが、珍しいイベントやシュート回りのクロスやポスト、落とすなどはアナウンスの流れを自然な文にするための繋ぎの語としてよく使われる。1シーンに同じイベントが含まれる場合はシーンの後半にあるものほど優先度が高いものとする。

表 4.6: イベントによる発話率

	発話数	全体数	発話率
ポスト	1	1	100.00%
落とす	4	4	100.00%
こぼれ球	4	4	100.00%
クロス	8	10	80.00%
コーナーキック	2	3	66.67%
クリア	2	3	66.67%
シュート	26	40	65.00%
パス	18	28	64.29%
フリーキック	5	10	50.00%
弾く	2	7	28.57%
ドリブル	3	14	21.43%

## 4.4 語彙変換

語彙変換は、アナウンス書き起こしデータより語彙リストを作成する。語彙にはパスやクロスが続いた場合、「から」、「そして」、「最後は」など連携した言い回しが存在し、これらはアナウンスを自然な文へと変換するために使用される。また、発話タイミングを調整するためにも多様されることがあり、図 4.5 では発話区間が余っているために演出として語を二回繰り返している。

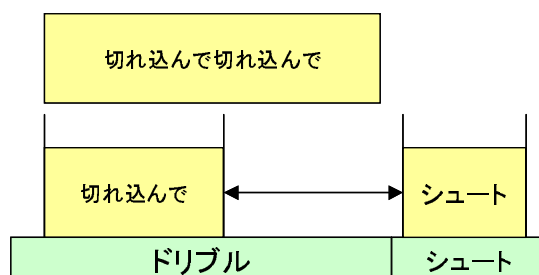


図 4.5: 語彙変換

語彙リストは以下表 4.7 の通りである .

表 4.7: 語彙リスト

インデクス	変換 1	変換 2	変換 3	変換 4
ポスト	ポスト	のポスト		
落とす	落とす	落として		
こぼれ球	こぼれ球	こぼれ球		
クロス	クロス	クロスに	そして	最後は
コーナーキック	コーナーキック	コーナーキックから		
クリア	クリア	クリアに	クリアボールを	
シュート	シュート	シュートに	のシュート	
パス	から	そして	最後は	
フリーキック	フリーキック	フリーキックから		
弾く	弾く	止める	止めるが	セーブ
ドリブル	ドリブル	持ち込んで	切れ込んで	選手名

語彙はイベントの連携パターン , また語彙の長さにより最適な語彙を選ぶ必要がある .

## 第5章 実装

### 5.1 プロセス

上記章の条件をシステムに組み込む．プロセスとしては以下のようなになる．

1. 最終イベントより攻めチームを決定
2. 初イベントより時間を決定
3. 発話タイミングの調整により前スロット&後スロットの位置を決定
4. 前スロット&後スロットにアナウンス挿入
5. シーン区間、発話タイミングを見ながら中スロットに優先度順にインデクス挿入  
(既存アナウンスの平均より発話スピードは 0.122 秒/モーラで計算)
6. イベントの関連性、余り時間を見ながら語彙変換

表 5.1 のようなインデクスがシーンに振られている場合，まず No.4 のイベントラベルより攻めチームを算出し、次に No.1 のイベントラベルより時間を算出、既存アナウンス音声の平均タイミングにて前スロットに挿入する．前スロットには「前半 3 分のオランダです。」というアナウンスが入る．また，最終スロットには，最後のイベントを行った選手の名前が入る．

次に中スロットに優先度の高い順に選手名とイベント名を挿入する．発話区間は前スロットと最終スロットの間の区間であり，またそこに「発話は実際のイベントよりも遅れて発話される」という条件が加わる．残り発話区間が無くなれば挿入を中止する．発話スピードは実際のアナウンスの平均スピードを 0.122 秒/モーラとし，計算した．挿入が終わると，インデクスを適切な語彙へと変換する．プロセスが終了すると，表 5.2 のようなアナウンス音声が作成される．

表 5.1: インデクス情報

No.	チーム	イベント	選手	開始時間	終了時間
1	オランダ	ドリブル	ロッペン	136.84	141.40
2	オランダ	クロス	ロッペン	141.40	142.20
3	オランダ	ドリブル	マカーイ	142.20	143.24
4	オランダ	シュート	マカーイ	143.24	144.72

表 5.2: 出力アナウンス

アナウンス	開始時間	終了時間
前半 3 分のオランダです。	137.68	139.14
ロッペン	139.25	139.74
持ち込んで	139.74	140.34
クロスに	141.40	141.89
マカーイ	142.20	142.69
シュート。	144.12	145.46

## 5.2 出力結果

以下、数例を示す。

### インデクス情報

表 5.3: インデクス情報サンプル 1

チーム	イベント	選手	開始時間	終了時間
デンマーク	パス	グラベセン	0.00	1.88
デンマーク	ポスト	サンド	1.88	3.20
デンマーク	シュート	トマソン	3.20	5.88



表 5.4: インデクス情報サンプル 2

チーム	イベント	選手	開始時間	終了時間
スウェーデン	ドリブル	バルヘルムション	55.28	59.36
スウェーデン	折り返す	バルヘルムション	59.36	60.18
デンマーク	こぼれ球	バルヘルムション	60.18	61.28
スウェーデン	シュート	ヨンソン	61.28	63.06

表 5.5: インデクス情報サンプル 3

チーム	イベント	選手	開始時間	終了時間
オランダ	ドリブル	セードルス	85.82	86.28
オランダ	パス	セードルス	86.28	87.78
オランダ	ドリブル	ファン・ニステローイ	87.78	88.86
オランダ	シュート	ファン・ニステローイ	88.86	89.28
ラトビア	弾く	コリンコ	89.28	91.30

表 5.6: インデクス情報サンプル 4

チーム	イベント	選手	開始時間	終了時間
オランダ	パス	セードルス	99.96	100.10
オランダ	ドリブル	レイツィハー	100.10	102.40
オランダ	ミドルシュート	レイツィハー	102.40	103.70
ラトビア	弾く	コリンコ	103.70	106.04

表 5.7: インデクス情報サンプル 5

チーム	イベント	選手	開始時間	終了時間
チェコ	パス	ガラセク	0.00	1.00
チェコ	ドリブル	ポポルスキ	1.00	2.50
ドイツ	ボールを奪う	シュナイダー	2.50	4.52
ドイツ	クロス	シュナイダー	4.52	5.70
ドイツ	落とす	シュバインタイガー	5.70	6.80
ドイツ	シュート	バラック	6.80	9.16

## 出力アナウンス

表 5.8: 出力アナウンスサンプル 1

アナウンス	開始時間	終了時間
前半1分のデンマークです。	0.84	2.42
グラベセン	2.53	3.14
サンド	3.14	3.50
のポスト	3.50	3.99
トマソン。	4.08	4.69

表 5.9: 出力アナウンスサンプル 2

アナウンス	開始時間	終了時間
前半1分のスウェーデンです。	56.12	57.82
バルヘルムション	57.93	58.91
持ち込んで	58.91	59.51
こぼれ球	60.18	60.67
ヨンソン。	62.16	62.77

表 5.10: 出力アナウンスサンプル 3

アナウンス	開始時間	終了時間
前半2分のオランダです。	86.66	88.12
ファン・ニステローイ	88.23	89.45
シュート。	89.74	90.35
コリンコ	90.35	90.83
弾く	90.83	91.08

表 5.11: 出力アナウンスサンプル4

アナウンス	開始時間	終了時間
前半2分のオランダです。	100.80	102.26
レイツィハー	102.37	103.10
ミドルシュート。	103.28	104.25
コリンコ	104.25	104.74
弾く	104.74	104.98

表 5.12: 出力アナウンスサンプル5

アナウンス	開始時間	終了時間
前半1分のドイツです。	0.84	2.18
シュナイダー	2.29	3.02
クロスに	4.52	5.01
シュバインタイガー	5.70	6.79
落として	6.79	7.28
バラック。	7.68	8.29

## タイミング対応図

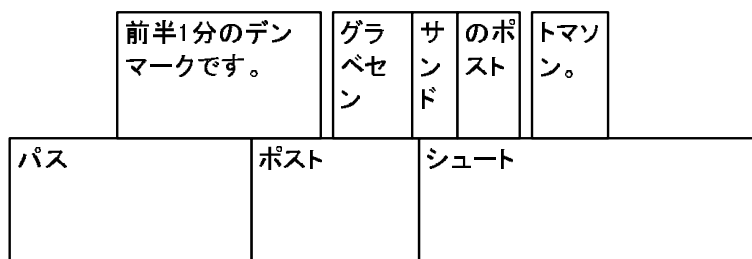


図 5.1: タイミング対応サンプル1

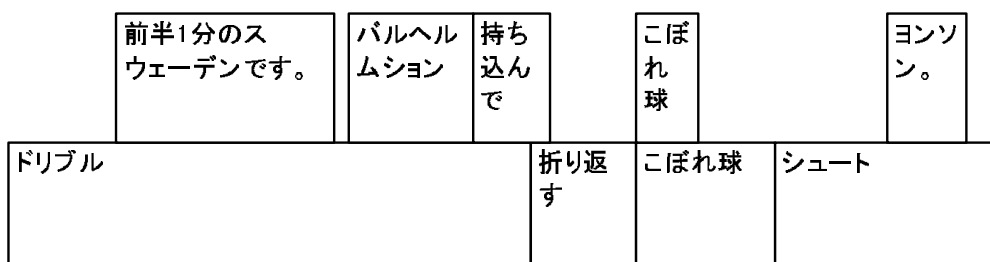


図 5.2: タイミング対応サンプル2

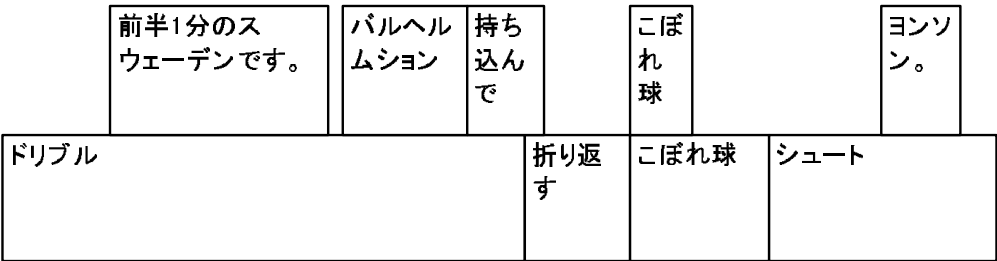


図 5.3: タイミング対応サンプル 3



図 5.4: タイミング対応サンプル 4

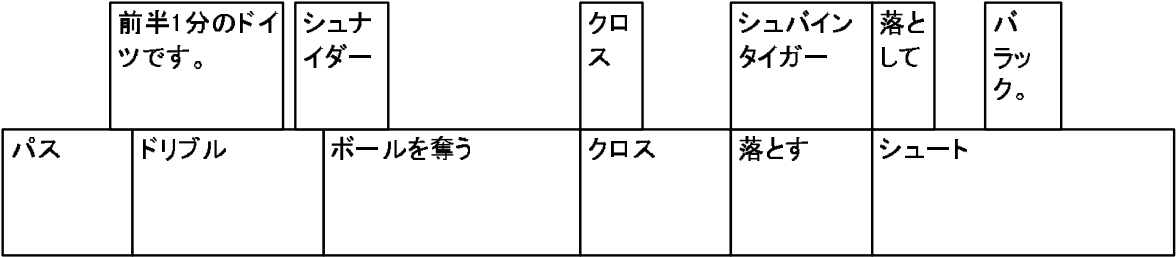


図 5.5: タイミング対応サンプル 5

## 第6章 評価

### 6.1 定量評価

作成したシステムの評価をする. 評価は既存アナウンス音声との挿入インデクスの一緻度にした. 以下表 6.1, 6.2 は出力した全 16 シーン中の誤り率, 正解率を示す.  $Correct$ (正解率) と  $Accuracy$ (正解精度) は次の式で定義される.

$$Correct = (\text{正解単語数}) / (\text{全単語数}) \quad (6.1)$$

$$Accuracy = ((\text{正解単語数}) \cup (\text{挿入誤り単語数})) / (\text{全単語数}) \quad (6.2)$$

ただし

$$(\text{全単語数}) = (\text{正解単語数}) + (\text{脱落誤り単語数}) \quad (6.3)$$

表 6.1: 誤り率

挿入誤り単語数	22	脱落誤り単語数	21
全作成単語数	88	全アナウンス単語数	87
挿入誤り率	25.0%	脱落誤り率	24.1%

表 6.2: 正解率

正解単語数	66
全単語数	87
Correct	75.9%
Accuracy	50.6%

## 第7章 考察

### 7.1 システム

発話タイミングの調整, 優先度による挿入インデックスの選択, 任意の語彙変換により自然なアナウンス文の出力が可能となった. しかし, システム全てのベースとなる発話スピードは既存のアナウンス音声の平均発話スピード 0.122 秒/モーラで計算している. 実際に発話は語彙や状況によりスピードが変化するため, スピードの変化を組み込んでいない現段階では実際のアナウンス音声とは挿入される語彙数が異なる可能性がある. この解決策として, 音声の代わりに文字による字幕スーパーで出力するという方法もあるが, 音声出力の方が試合を盛り上げる演出効果は高い.

### 7.2 評価

今回は定量評価を行った. その結果 Correct(正解率) は 75.9%となったが, Accuracy(正解精度) は 50.6%と低い結果になった. これは発話スピードの問題と共に, 実際のアナウンサーが毎回アナウンスする内容には元々ある程度の誤差が含まれるという原因が考えられる. また, 形容詞や接続詞といった文を加工する表現を今回挿入していないため, その分の脱落誤りが増えたということも考えられる. しかし, この評価を主観評価にすれば, 既存アナウンス音声が必ずしも正解とは限らない. 本研究の目的は映像の内容の理解しやすさ, 映像の流れの理解しやすさを高めるためのアナウンス音声の作成である. 主観評価を行うためには音声出力ツールが必要となるが, 合成音声を使用した場合, 同じ抑揚のまま作成した発話時間に納まるよう発話スピードを変更するため, 不自然な音声になる可能性がある. したがって今回は定量評価のみとした. 今後音声の抑揚を踏まえた研究と組み合わせることにより主観評価が可能なものとなる.

## 第8章 まとめ・今後の課題

### 8.1 まとめ

本稿ではサッカー要約映像におけるインデクス情報を利用したアナウンス音声の自動生成を目標とし、要約映像作成時におけるインデクス情報を再利用し、既存アナウンス音声の生成ルールを分析することにより、優先度の高いインデクスを順に発話タイミングを考慮しながら任意の語彙に変換して挿入していくシステムを作った。これによりユーザの映像の内容の理解しやすさ、映像の流れの理解しやすさを高める。

しかし、現在出力しているのはアナウンス内容と発話時間のみである。発話スピードは実際のアナウンスの平均スピードで計算しているが、実際の音声は語彙や状況により発話スピードが変化するため、音声出力する場合には発話スピードの強弱を検討する必要がある。また、これを合成音声による出力にすると、試合展開と著しくそぐわない抑揚で発話される可能性があり、抑揚についても検討する必要がある。

### 8.2 今後の課題

今後の研究課題は以下の3項目である。

- 形容詞, 接続詞の挿入
- 抑揚を考慮した音声出力
- 主観評価 & システムの改良

本研究で作成したアナウンス音声では1シーンを独立したものと考え、シーン間での繋がりが存在しない。しかし、形容詞, 接続詞表現をアナウンス音声に加えることにより、より映像の内容把握を補助するアナウンス音声の作成が可能となる [8]。例としては「しか

し」や「そして」のような逆接や順接をシーンの冒頭に付加することにより、そのシーンの展開が事前に理解可能となる。また「今度は」や「また」などの前シーンの履歴を踏まえた情報を付加することにより、映像を盛り上げる演出効果を発揮する。

音声出力には音声の抑揚を考慮する必要がある。抑揚を踏まえた上での音声出力が可能となれば、本研究にて作成したシステムと組み合わせることによりアナウンス音声の自動生成が可能となる。

最後に主観評価、システムの改良を繰り返すことにより、より完成されたシステムが完成する。



## 関連図書

- [1] 益満健, 越後富夫, “映像重要度を用いたパーソナライズ要約映像作成手法”, 信学論 (D-II), vol.J84-D-II, no.8, pp.1848-1855, Aug. 2001
- [2] 河合吉彦, 馬場口登, 北橋忠宏, “個人適応を指向したスポーツ要約映像の生成法”, 信学技報, PRMU2000-171, pp.83-90, Jan.2001.
- [3] 丸尾二郎, 岩井儀雄, 谷内田正彦, 越後富夫, 飯作俊一, “サッカー映像からの特定映像イベントの抽出”, 信学技報, PRMU99-41, pp.31-38, 1997-07.
- [4] 宮内進吾, 馬場口登, 北橋忠宏, “音声・言語・映像の協調的処理によるスポーツ映像からのイベント検出”, 信学技報, PRMU2001-214, pp.63-70, 2002-01.
- [5] 稲葉大樹, “フィールド情報に基づくサッカー近景映像の自動インデクシング”, 白井研究室修士論文, 2003
- [6] 塩崎崇, “音響信号処理に基づくサッカー映像のインデクシング手法”, 白井研究室卒業論文, 2003
- [7] 河合吉彦, 馬場口登, 北橋忠, “放送型スポーツ映像におけるデジタルビデオ効果に着目したリプレイシーンの検出の一手法”, 信学論 2001/2 Vol.J84-D-2 No.2, pp.432-435, 2001-02
- [8] 白田由香利, 橋本隆子, 飯沢篤志, “ダイジェスト説明文生成方式についての検討”, 情報処理学会, DBS120-15, 2000.1.24

# 謝辞

本研究を行うにあたり、研究環境を整えて頂き、適切な御指導をいただいた白井克彦教授に感謝いたします。

本研究の方向性を定め、御多忙にも拘らずにいらして御指導していただいた早稲田大学スポーツ科学部教授の誉田先生、東洋大学講師の村上先生には大変感謝しております。御二方のおかげで本論分を完成させることが出来ました。どうも有難う御座いました。

サッカー要約映像の自動作成という一大プロジェクトを立ち上げ、私たちに遣り甲斐のあるテーマを残してくださった先輩の稲葉大樹氏にも感謝しております。まさかこのプロジェクトが3年越しのものとなるとは思いませんでした。お陰様で同チームの後輩も出来、楽しい研究室ライフを送ることが出来ました。有難う御座いました。

M1 の加賀明久君、鈴木健司君、塩崎崇君、先輩としてあまり役に立てなくてごめんなさい。私の方は君たちがいてくれて何度も救われることがありました。加賀君、毎回のゼミ連絡ご苦労様でした。鈴木君、君が一番研究に対してやる気がありました。今までのプロジェクトに拘らず自分のやりたいと思う研究をしてください。塩崎君、君は頭がいい子です。その才能を無駄遣いさせちゃっているようで申し訳なかったです。

その他、研究分野が違うにも関わらず、毎度適切な助言を与えてくださったドクターの谷口徹氏にもお礼を言わせて頂きます。有難う御座いました。音声部屋、画像部屋と離れた環境の中で最もこちらの飲み会に積極的に参加してくれたのが谷口氏だと思います。いろいろとご迷惑をかけたこともあるでしょうが、私たちのような駄目な画像 M2 と付き合ってください。有難う御座いました。

また、別チームの後輩の板野貴之君、石毛克尚君、松井和教君にも大変お世話になりました。あなたたちはとても愉快でした。

画像部屋の同期である長船大毅氏、川口克則氏、鈴木康郎氏、森野岳宏氏にも感謝しております。あなたたちと共に M2 になれて、得たものと失ったものがどっこいどっこいです。

謝辞

最後に、経済面・生活面にわたって多大な援助をしていただき、貴重な研究活動の機会を与えてくださった両親に心から感謝いたします。

2005 年 2 月

平林 衛